Students' views of oral performance assessment in mathematics: straddling the 'assessment of' and 'assessment for' learning divide

Paola Iannone

School of Education and Lifelong Learning, University of East Anglia, UK

Adrian Simpson

School of Education, Durham University, UK

Corresponding Author: Paola Iannone, School of Education and Lifelong Learning, University of East Anglia, Norwich, NR4 7TJ, UK. Email: p.iannone@uea.ac.uk.

Paola Iannone is Senior Lecturer in Mathematics Education at the University of East Anglia.

Adrian Simpson is The Principal of Josephine Butler College and Reader in Mathematics Education at Durham University.

Acknowledgements

This work was supported by the MSOR Network in the UK through the Mathematical Sciences HE Curriculum Innovation Project, part of the National HE STEM Programme.

Students' views of oral performance assessment in mathematics: straddling the 'assessment of' and 'assessment for' learning divide

This paper explores the views of a group of students who took an oral performance assessment in a first year mathematics module. Such assessments are unusual for most subjects in the UK, but particularly within the generally homogenous assessment diet of undergraduate mathematics. The evidence presented here resonates with some, but not all, of the existing literature on oral assessment and suggests that, despite concerns about anxiety and fairness, students see oral assessments as encouraging a focus on understanding, being relatively authentic and reactive to their needs. We argue that, suitably implemented, oral assessment may be a viable assessment method for straddling the 'assessment for' and 'assessment of' learning divide in higher education.

Keywords: assessment for learning; assessment of learning; oral assessment; mathematics.

Introduction

Given the pressure to diversify assessment methods, while maintaining the confidence of both the academic community and the student body, it is important that we evaluate the implementation of new methods carefully and in an appropriate context. Brown and Knight (1994) argued that using a diversity of methods is important in avoiding disadvantaging students and Brown, Bull and Pendlebury (2013) suggested that homogenous assessment systems lead to reproductive learning styles, while systems with a variety of assessment methods are linked to more flexible thinkers. Birenbaum et al. (2006) argued that change is important not just in terms of the assessment methods used, but also the purposes they are designed to serve.

However, some have questioned whether sufficient attention has been placed on context when recommendations are made about modifying patterns of teaching and assessment. Joughin (2010) noted that some key influences on higher education policy and practice may have been over-generalized and that research involving students in

particular countries, universities and disciplines may not easily generalize to other contexts. For example, Iannone and Simpson (2014) noted that the vast majority of participants in studies on students' assessment preferences came from just five subject areas (psychology, biology, economics, education and engineering) with none apparently coming from what Biglan (1973) called the 'hard, pure' sciences. Norton, Norton and Shannon (2013) noted that little work has focused on the effect of discipline on assessment practices, though Yorke (2011) argued for a need for a better understanding of the mix of assessment practices in different disciplines and its impact on grades and degree classifications.

In the hard-pure sciences, and particularly the context of mathematics, there have also been repeated calls for increasing the variety of assessment methods (see, for example, Steen [2006]). Yet, the pattern, in the UK at least, remains remarkably uniform, with a heavy emphasis on the closed book examination. A recent survey of assessment practices showed that the median contribution to the final degree classification from closed book examinations was over 70% (Iannone and Simpson 2012a). Within what they called the 'assessment diet', the survey authors found that final year projects, weekly exercise sheets and computing projects were common alternatives in making up the remaining 30%.

This uniform diet may reflect a key concern of the mathematics community that the nature of the subject requires particular forms of teaching and assessment which may not necessarily fit well with institutional pressures to implement particular methods across disciplines. For example, the London Mathematical Society (the largest learned society for mathematics in the UK) raised concerns about whether mapping of mathematics examination marks on to 'a university scale, defined by qualitative descriptors' is turning poor performances into passing grades (LMS 2010, p1).

Moreover, students appear to share this sense of the particular nature of their subject: Iannone and Simpson (2014) found that, contrary to the findings from more generalist literature, the mathematics students in their sample tended to prefer to be assessed by traditional methods and perceived them as better at discriminating between students on the grounds of ability.

The views of students are therefore very important in evaluating new assessment methods: by understanding their perspective, we can learn about the impact of the assessment on their learning, issues related to its implementation and the place of the assessment method within the wider assessment diet.

Oral Performance Assessment

Oral assessment has tended to be overlooked in research into assessment in higher education, despite being used widely in many countries for most academic subjects (De Vita and Case 2003). Joughin (1998) provided a comprehensive categorization system for oral assessments based on a number of different dimensions. These included

- whether the aim of the assessment is to assess knowledge and understanding, applied problem solving ability, interpersonal skills or personal qualities;
- whether the action is presentation, dialogue or some combination;
- whether the structure is closed (with fixed questions from which one cannot deviate), open (where both student and assessor are free to deviate at will) or somewhere in between;
- whether it is authentic (simulating a real world situation) or de-contextualised and
- whether the medium used is purely oral or whether other media are used.

So what one calls an ‘oral assessment’ can vary considerably.

Our study was based on what we call an ‘oral performance assessment’. It involves assessing knowledge and understanding with elements of both presentation and dialogue, it is somewhat decontextualized (though the issue of ‘authenticity’ arose, as noted later), has a relatively open structure and combines oral medium with writing on a board.

Given the wide number of forms an oral assessment may take, it is important to interpret the evidence in the literature in the context of the particular implementations on which it reports to evaluate its applicability.

Some of this literature points to issues of anxiety. For example, Diaz, Glass, Arnkoff and Tanofsky-Kraff (2001) noted that state anxiety was somewhat predictive of oral argument score for first year law students at a private US institution. However, this was a form of public argument, mirroring arguing a case in court in a large group setting, and so fits only one combination of Joughin’s dimensions. Moreover, the study was not a comparison between assessment methods. Comparing methods may indicate that the issue of anxiety is more nuanced: for example, Sparfeldt, Rost, Baumeister and Christ (2013) showed similar levels of test anxiety (albeit in different forms) in German students between oral and written assessment conditions for mathematics and the German language.

Henderson, Lloyd and Scott (2002) looked at oral assessments in social work education and found that increased experience with this method alleviated initial anxiety: by the time the students had entered professional practice, they recognised the importance of oral assessments in terms of authenticity and value to employment. Huxham, Campbell and Westwood (2012) also noted that anxiety might be offset by value and authenticity in their research with biology undergraduates and questioned

whether the anxiety was inherent in the orality of the assessment or whether it was tied to its novelty.

A second area of concern with oral modes of assessment is fairness. Heyneman (2004) raised concerns that, because oral assessment is normally carried out in private, it is prone to corruption. It is also believed to be prone to potential bias (whether conscious or unconscious) because oral performance assessment cannot be anonymised (Davis and Karunathilake 2005). Again, however, it is not clear how significant this is in comparison to other assessment methods. It can be difficult to completely remove non-essential factors from any assessment: for example, Briggs (1980) suggested that even handwriting style could have a significant effect on marks for identical essays, marked anonymously.

Other authors have suggested positive aspects of oral assessments in relation to fairness. Joughin (1998) noted that oral assessment all but eliminates the possibility of plagiarism and other cheating and, unlike examinations, oral assessment with a tutor prevents one small gap in knowledge stalling a solution. This is particularly true in subjects like mathematics when failing to answer the first part of a traditional examination question correctly often results in the inability to answer the subsequent parts. Moreover, De Vita and Case (2003) suggested that oral performance assessment facilitates a more legitimate assessment of non-native-speaking students who are often more proficient in speaking than writing in their second language, and can allow well-trained assessors to differentiate language difficulties from conceptual ones.

The aims for this study were to uncover and describe the views of mathematics students who took part in a small trial of oral performance assessment. As an exploratory study, it also investigated the extent to which those views resonate with those expressed in the literature.

Implementation

Agreement was obtained to introduce an element of oral performance assessment within a first year module in discrete mathematics at a research intensive UK university. Ethical approval for the accompanying research study was also obtained. The module covered introductory combinatorics, algorithms, graph theory and optimization. It was normally assessed by final closed book examination (worth 90% of the final module mark) and coursework (worth 10%) consisting of 10 weekly exercise sheets. The module was normally taught through three one-hour lectures and one one-hour tutorial (generally in groups of around 20 students and focused on the set exercises) each week. This pattern of assessment is typical for first year modules at that university (and it is a common pattern across many similar mathematics departments in the UK).

There were 108 students enrolled in the course and the oral performance assessment (which was called a 'one-to-one tutorial' in communications with the students) replaced one of the weekly tutorials.

The assessment for that week was based on four questions the lecturer had planned for the usual coursework system (table 1).

[Tab 1 goes here]

The students were each assigned a ten-minute slot (normally within their usual tutorial time) to attend their one-to-one tutorial. The assessments were conducted individually by one of five different assessors: the lecturer, the usual tutors and two other experienced mathematics lecturers (the authors of this paper). To ensure that the assessments could be moderated or marks challenged if necessary, they were video recorded.

Given the novelty of the method, various measures were used to ensure consistency of process and marking. The five assessors met to develop an agreed marking matrix (shown in figure 2), protocols and contingent questioning areas. The students were told that they would only be asked to solve two of the four questions: they could choose one and the assessor would choose one remaining question randomly (by, for example, flipping a coin). In fact, because it was deemed that questions A and B assessed similar mathematical processes (producing a formal written proof) and questions C and D were also similar in requiring the statement of an answer with a less formal justification, it was agreed that if the students chose a question from the first pair, the assessor would randomly choose one of the second pair (or vice versa). The mark (out of ten, as with weekly written exercise sheets) was noted immediately on completion of the tutorial and then emailed to the students a few days after the assessment.

[Tab 2 goes here]

Contingent questioning and dialogue is an important feature of this form of oral assessment (Joughin 1998), allowing the tutor to explore the student's understanding based on their on-going performance in the tutorial. It might involve drawing the student's attention to a problematic issue with their answer, explore their understanding of the source of mathematics they are using, examine the use of the premises of the question in their solution etc. For example, contingent questions suggested by the assessors included asking students where the number 11 is used explicitly in their solution to question A or how they might go about drawing forests of 6 vertices for question D. While such questioning needs to be individual and spontaneous, the

assessors felt they could predict a number of areas which might arise and agreed some potential questions beforehand.

Recognising Huxham et al's (2012) concern that oral assessment anxiety may be primarily related to its unfamiliarity, a number of measures were taken to prepare students. A short talk about the assessment was given during one of the lectures at which students could ask questions. Students were also emailed with an outline of the format of the assessment and further encouraged to ask questions. All but one of the questions asked related to the practicalities of the process, such as how they would find out about the time of their tutorial, how long they were expected to last etc. Only one student raised concerns about this form of assessment and this related to her phobia of being video recorded; in this case, special permission was given for her to have two people present for the assessment instead of being videoed.

Of the 108 students on the module, 99 attended the oral assessment. It is common that a small number of students will fail to submit any work for a weekly written coursework task or attend their weekly seminar, so the failure of 9 students to attend was neither worrying nor disproportionate.

Data Gathering and Analysis

Some of the analysis of the outcomes of the oral performance assessment, including the outcome grades and responses to an assessment experience questionnaire is detailed in Iannone and Simpson (2012b). The focus here is on the students' views of the assessment and uncovering key themes within those views.

After the end of the oral assessments, students were contacted by email to take part in an interview study. Nineteen students (nine female and ten male) agreed. The interviews took place one week after the oral assessments and were conducted one-to-one in an office (with the first author of the paper as the interviewer). Inevitably, the

students who chose to take part in the interviews might not represent the views of the whole group equally, but we note below that they do give a wide range of opinions on the assessment methods.

The interviews were semi-structured. That is, the interviewer had a number of core issues to address and which they would introduce if the interviewee did not spontaneously address them. The interviews otherwise took the form of a focused conversation which allowed the students' views to emerge (Gillham 2005). The core issues included their experiences of different assessment methods, their impressions of the one-to-one tutorial and their views of the advantages and disadvantages of the assessment method. The focused nature of the interviews meant they lasted around 20 minutes; they were audio recorded and subsequently fully transcribed.

The analysis of the interviews followed the inductive pattern established by the thematic networks approach (Attride-Stirling 2001) which facilitates uncovering themes within the transcripts and the organising those themes into a framework. Both authors independently coded a small number of transcripts and developed initial basic themes. These were compared and negotiated and an inductive cycle of developing themes and analysing transcripts took place before the final organizing themes were agreed.

Emergent Themes

Figure 3 shows the resulting thematic network, which has three main organizing themes – *anxiety*, *fairness* and *understanding* – which each had further subthemes, and two further minor themes: *reactiveness* and *authenticity*.

[Fig 3 goes here]

Figure 3 illustrates that anxiety and fairness can be seen as themes related to the implementation of the assessment or the process, while understanding, reactiveness and authenticity might be seen as themes concerned with the properties or characteristics of the assessment, as experienced by the students.

Process – Anxiety

All students interviewed talked about anxiety in relation to the oral assessment. The literature often highlights the levels of anxiety which might be associated with this mode of assessment. However some literature also suggests this may be related to its novelty, rather than intrinsic to the assessment method. Indeed, one of the most common comments was about the lack of familiarity:

> I think that you know what to expect from an exam paper … You don't know … [Be]cause obviously you don't know who is gonna be your tutor. You don't know how it's gonna start …I mean to be fair it was pretty much exactly what we were told it was going to be. You prepared your questions. You pick one and get made to do another one. So it wasn't any nasty surprises. But there is a bit of a weird concept.
> (Ruby)

Ruby (all students have been given pseudonyms) seems to be suggesting that the requirements of examination papers are highly predictable, but that the lack of experience with oral performance assessments requires a step into the unknown, despite the effort put in to preparing the students for this experience, which she appears to acknowledge. However, Ruby's comment also highlights the idea that it is the expectation of the assessment which may cause the anxiety, and the reality may be somewhat different. This also appears in Arthur's succinct comment:

> Em, I felt a bit worried. I felt worried... but afterwards it was ok.
> (Arthur)

Not all of the anxiety, though, may have been rooted in lack of familiarity. Some students mentioned a sense of being exposed:

> … because all the way in school you haven't done anything like that – you have just done written examinations so… maybe because you see a member of staff you think oh they are going to…. they will look down on you if you don't know the answers …
> (Hollis)

Of course, all summative assessment methods have the potential to expose the student in the way Hollis describes. Even the weekly exercise sheets, which the students handed in for each of the other weeks of the module, would have shown their tutors some of their misunderstandings. However, there is a perception here that the one-to-one nature of the oral assessment may make that sense of exposure all the more immediate (an issue which is also addressed below).

Having noted that, some saw the level of anxiety as having a potentially positive effect, appearing to indicate that the extra pressure led to more effort in preparing for the assessment:

> … with just the piece of normal homework there is less pressure... you don't feel like you're going to have to explain more because... I don't know, all people are different, but for me, I always feel like I want to do the best I can, especially when someone is staring at you, when you're presenting your knowledge and your work.
> (Angelina)

Angelina seems to echo Hollis's concern about exposure, but in a more positive way. The possibility of having to explain the material 'live' seems to encourage an emphasis on understanding and, as with Arthur, Angelina also suggested the novelty had an impact on the anxiety:

> First, I was a bit worried about it but after it, it is definitely better.

(Angelina)

Process – Fairness

The issue of fairness, in one guise or another, was also discussed in almost all of the interviews. The most common comments focused on the number of different assessors, the contingent questions and the comparability between the assessment as experienced by different students.

> . . . because there was like two or three tutors who were doing the tutorials and they were . . . I think they were run like quite differently. For some people said that they were asked for some like …extension questions whereas I just answered the question out of the ones I prepared.
> (Delia)

Delia appears to be suggesting that the agreed procedure for the assessments was different between tutors, which an examination of the videos suggests was not the case. However, her discussion with her peers has given her a sense that she did not recognize the time spent on contingent questions in her tutorial as comparable to others, and clearly contingent questioning was dependent on individual performance.

As noted above, however, contingent questioning is a crucial feature of the oral performance assessment, but some saw this as introducing unfairness or randomness:

> ... also I talked to my friends about how theirs went… although we did the same question a lot of us talked about completely different things. So it was very hard to like make sure someone is guided on the same way as everyone else. I think it would be a problem…
> (Tom)

But what Tom sees as a problem, others saw as a benefit.

> Oh no, I don't think it is a problem – ... because at the end of the day, I guess that the whole point of the one to one [*tutorial*] is your thinking…. What your problems

> are … kind of like the main goal – you want to solve your problems … if every tutor was doing exactly the same thing in a one to one tutorial you may as well have a class with 15 people doing all the same thing. If it is a one to one tutorial I think it is most useful if it is unique to what your needs are or what your struggle is so… I don't think that having different people doing it is a problem at all…
> (Rick)

> It can't be uniform in the same way that everyone gets the same questions can it? You know when you get helped… how can you decide how to … because if someone falls down on different questions how can you say – well – this amount of help is the same as this… and it gets a bit hard to regulate.
> (Corrina)

Rick, in contrast to Tom, seems to acknowledge that an assessment focused on exploring someone's understanding inevitably has to be individual. Corrina's focus is on the tutorial as a teaching and learning opportunity as much as a summative assessment: it is a time when the student can be helped and, since people make mistakes at different places, or have different misunderstandings, of necessity, the help has to be individual.

So there was no consensus in relation to fairness as it arose in relation to different tutors and questions. However, many of the issues about fairness in the literature did not appear as such in the interviews. Some students showed for example an awareness of issues such as bias, but felt that fairness in this sense was guaranteed by the videos:

> . . . also with written work, we hand in homework and they are corrected by different people so it can't be marked in exactly the same way. It is the same with an oral exam. But the fact that it has been recorded guarantees that . . . if we don't agree with something we get ... we can go back to it.
> (Isis)

> I agree with that [*using videos to moderate*]… you don't want just one person… they may see you as someone who doesn't understand but then someone else may see you as one who does.
> (Hollis)

Isis also notes that the issue of different assessors does affect other assessment methods (such as weekly exercise sheets which are assigned to different tutors). But the main issue for several students, as exemplified by Isis and Hollis, is that while bias and unfair marking could occur, the relatively simple expedient of recording them goes a long way to dealing with the problem.

Characteristics – Understanding

All of the students referred to the impact of the assessment on understanding, in terms of the understanding they developed in preparing for the tutorial, its development through the interaction in the tutorial or the way in which that understanding was evaluated.

Many students talked about the difference in the way they prepared for their assessment and they way they normally prepared for their weekly exercise sheets and the associated tutorials. We have already heard an element of this from Angelina in discussing the positive side of the pressure felt by the oral assessment. But she later spoke to the issue directly, as did others:

> So I felt each question you have to completely understand and be able to explain it to anyone . . .
> (Marie)

> …and to prepare for that [*the oral performance assessment*] you do have to understand what you're doing. Well not only the problems in themselves, like the general case.
> (Angelina)

> I prepared a lot more for the oral one obviously [be]cause I needed that I have to … generally understand what I was talking about so I can actually communicate it whereas I think that sometimes like on paper you just kind of… write the answer and not necessarily completely understand the whole theory behind it. So, I think I definitely had to do more work to do this.
> (Tom)

Marie, Angelina and Tom all appear to indicate that the extra work is focused not simply on developing fluency, but on understanding. For Marie, the source of this need is in the expectation that she will be asked to explain, not simply reproduce. In Angelina's terms, however, they don't simply have to be able to solve the problem, but must see where it fits as an instance of a more general case. Tom includes both factors.

This highlights again the importance of the contingent questioning (which was seen both positively and negatively in relation to fairness): knowing that they could be asked questions which probe understanding or ask them how they might start solving a slightly different question seems to lead students to engage differently with preparation. Marie is explicit about this:

> …with homework, you're sitting in your room... most the time you switch off or you would switch off for the majority of it, but there will be bits where you write it down, and you'll look back at it later, and you will go: why did I write this down? Whereas when you're put under the pressure of someone is looking at you, you've got to make sure that you're doing it correctly. So, you look through it. [...] .. yeah, it was a lot more useful than normally.
> (Marie)

but note that, as with the issue in the anxiety theme, there is an implicit sense of the possibility of being exposed with 'the pressure of someone … looking at you'.

Even for someone who felt that their preparation was quite similar to that for the weekly written exercises, there are some interesting differences:

> I am not great at mental maths so I prefer to write my thoughts whereas I think with the oral like assessment I found like I tried to jump steps because I didn't want to do the 'low' steps I would usually do. That's the only difference for me [be]cause I like to write things out quite thoroughly - that's how I have always done … how I ever did my maths
> (Ruby)

Ruby seems to have been influenced by the assessment to focus more broadly on the key steps and away from the minutiae. Presumably, she felt that she would be able to reconstruct the 'low steps' and this allowed her to concentrate on more conceptual material.

Interestingly, a small number of students saw the oral assessment preparation as moving them away from understanding and towards memorizing, possibly as they were told they were not allowed to consult notes or materials during the one-to-one tutorial:

> … when I do the homework usually.. I do the homework like the answer to the questions … but we couldn't take the paper so I had to memorise like the answer to some extent so I could… I think I need to be able to remember the answers to this specific questions better to be able to present them to someone like… the specific answer to the question usually I sort of get an understanding of the subject and then you could perhaps answer a similar question in the exam but I thought that when you went in you needed to quite quickly know … you just had to memorise it.
> (Corrina)

Corrina seems to have approached the task in a similar way to the weekly written exercises and then added a stage of memorizing. When she uses the word 'understanding' it appears to be an approach to learning focused on answering a similar question in a written examination with, presumably, the application of a well rehearsed approach.

However, most students, as those above, seem to have been talking about using preparation in the tutorial to try to develop a less instrumental form of understanding (in

the sense of Skemp [1976]). But they also noted that some of the understanding might develop within the tutorial itself.

> … when you explain it to someone else, you sort of instinctively understand it better yourself because you're having to know it quite inside out in order to explain it to someone else…. You almost subconsciously hear it back so… So, you understand it quite a lot.
> (James)

James seems to suggest that the act of explaining itself can aid understanding. He gives the sense of his understanding having the form of a good, but rather loose, network of connections which gets tightened up and solidified through the act of presenting.

In addition to explaining helping to strengthen understanding, the one-to-one nature can allow for both diagnosis and remedy for misunderstandings:

> I think it helped in a sense that when you're asked why and to explain things, you realise that you don't actually fully understand and maybe that you need to go and look over or whatever… because you've been asked and you can't explain it…
> (David)

> …it made me think of it and I thought it was really good to discuss it immediately because even if we had some problems in understanding and stuff we could… it was an opportunity to ask.
> (Isis)

David notes that explaining can expose misunderstandings, as can being asked direct questions and Isis saw the tutorial as much as a teaching situation as an assessment. While it is not clear that everyone felt the same, clearly she felt able to ask when she didn't understand something.

The final basic theme in relation to understanding was the role of the oral performance assessment as an evaluation of understanding and the way it afforded

tutors different ways of recognizing it. In some cases, this was phrased as a comparison to the weekly exercise sheets:

> I think that they [*the assessors*] get to know your ability a lot easier than just a written piece of work because you're seeing them make sense of the problem in front of you rather than just once they have done it. Because when I do my homework … my own thought processes are like messy and then I write up my homework. So obviously they won't see what I have gone through or what mistakes I've made. So you can see that a bit more with the assessment.
> (Ruby)

> …with the written work at home … – you could just go on the internet or do work with other people whereas if you are asked directly to solve problems, even if you have prepared them in advance it is clear if you have understood the topic or not, I think.
> (Isis)

> … as well as just getting the basic questions that you get in the homework, then you do the answers in them perhaps, but if there is something that isn't quite clear that you understand it or not, the assessor will not be able to ask you more questions around it. … [*in the tutorials]* it wasn't just based on the actual question you're giving. They could see... They could spot where you're, well, missing the understandings... and then... like ask you the questions around just to see if you just do the homework questions. …[*In weekly written exercises, they*] can only assess you on the actual questions that they're giving so they can't actually see what you understand in the full context around it.
> (Alberta)

This sense of the ease with which the assessors can evaluate understanding was a near universal theme. In the quotations from Isis and Alberta, as well as from many other students, the implication is that one of the key tools the assessors use is individualized and contingent questioning. Coupled with the theme of fairness, particularly in relation to the comparability of the questions, we might see a tension here. On the one hand the students see the oral performance assessments as facilitating

the evaluation of their understanding, but on the other hand, to do so the assessors need to ask questions which are dependent on the individual's presentation of the problem and previous responses.

Characteristics – Reactiveness

One of the characteristics repeatedly highlighted by the students was the reactiveness of the situation: that is, the assessor could respond to the performance of the student and to their needs. While some contingent questioning had been planned in to the process, the students noted a level of assessors' reactiveness which goes beyond this.

One particular type of reactiveness was the opportunity for the assessor to help the student over a gap in knowledge, small slip or other impasse:

> In an exam, if you don't know a question, you can get stuck and you can't go any further with the question but if you get a little bit of guidance maybe you can show that you can work through a bit more. You've got a second chance.
> (Arthur)

> The tutorial, I think it was more targeted, a lot more targeted because… I felt more focused around me and my thought processes rather than in the tutorial where it is focused around the questions. And a lot of [*standard*] tutorials might give you time to think about it and ask how to do it but it feels like it is one set, kind of logical answers. You know they [*tutorial leaders*] give their answers rather than looking at where your weaknesses are and what bits you need to understand
> (Ruby)

This clearly relates to the issue above about the assessment of understanding, but the emphasis here is on the individuality of the assessment experience. With Ruby, the notion of 'targeting' her personal thought processes and providing reactive advice about what she needs to work on to improve her performance appears to contrast with her view of weekly exercise sheets and their associated tutorials.

The reactiveness also shows itself in the immediacy of the feedback obtained. While some students were concerned that they received no formal, written feedback, many valued the immediate verbal feedback in the tutorial:

> … because the person is listening to you as well and you know if you're wrong, you'll pick it up almost immediately because you'll know… you know what I'm talking about. So, I think that's really important. … [*The weekly exercise sheet*] was pretty much, I just wrote it down on paper and we didn't get it back till two months later or something so. And the feedback was just really short. It wasn't very detailed. … So I feel like cause when you're talking to someone at least, my feedback were almost pretty much be immediate. [Be]cause you know what I'm doing.
> (Sarah)

The implication from Sarah is that written weekly exercises do tend to have formal feedback, but too often it is lacking detail and too late to be of much practical use (though 'two months' is probably an exaggeration). The immediacy of the feedback in the oral assessment is something she and many other students valued.

Characteristics – Authenticity

Joughin (1998) lists 'authenticity' as one of the dimensions in his analysis of oral assessments. He describes it as 'the extent to which assessment replicates the context of professional practice or "real life" '(p371). In this case, many students talked about the extent to which the tasks of communicating their solutions and responding to questions might reflect skills required in the job market.

> I think it is very important because when you go out in the wider world you are not going to be … you are not going to get somebody saying to you – oh do this problem and hand it to me next week. Somebody is going to say to you: oh can you sort this out and can you explain to me how to do it. And I think it is a much more useful skill to be able to explain your thinking to somebody else than to be able to

> write the answer and hand it in – I think it is a much more useful skill …
> (Rick)

Rick ties the form of the oral performance assessment to his perception of the reality of employment and negatively compares that with the weekly exercise sheets. However, the basic theme of authenticity developed from the interviews isn't just restricted to the students' perceptions of 'real life' or professional practice, it also encompasses the development of skills needed later in their course:

> Especially seeing here you have a project in third year, you have to do an oral presentation and to suddenly be told in the third year that you have to do an oral presentation and haven't done public speaking before so. … That kind of thing doesn't worry me: it is the fact that nothing mathematical I have ever done before was in an oral environment. I mean, you can argue a debate for instance, it is completely different and you need to communicate your ideas but like… in an oral manner. So I just think it would be invaluable especially… even if you have one assessment a year on one subject just because I need something… It needs to be part of… For the sake of the argument it could be formative but then I don't know how constructive it would be.
> (Marie)

Marie makes the tangential, but often repeated, point at the end of this quotation that some students are less motivated to engage with assessment which is not seen as contributing marks. However, her main point is about what we might call 'internal authenticity': the extent to which the assessment replicates or prepares students for other assessment contexts from within later parts of the course.

As at other universities in the UK, this mathematics department has a project module, which contributes a considerable proportion of the final year mark and is partially based on a presentation and question-and-answer session. It would appear that Marie, among others, has concerns that such a high stakes assessment is based on skills which have not been developed or assessed earlier. It is telling that, as a first year

student, Marie is already concerned about the project and preparation for it, even though it is two or more years away.

Many students held this same view that the oral performance assessment, particularly if repeated, would help develop mathematical communication skills which might be needed later in the course.

Discussion

It is important that alternative forms of assessment are evaluated and part of any evaluation should take into account the views of students. The oral performance assessment implemented here was a relatively small trial of an alternative to the generally homogenous assessment diet of closed book examinations, leavened by weekly exercise sheets and occasional projects. Many of the students taking part in the study talked incidentally about examinations as the key form of assessment, even though for the most part their focus was on the oral performance assessment. They tended to compare it to the weekly exercise sheets which this intervention intended to replace.

There was some resonance between students' views and the literature. Certainly both anxiety and fairness arose as organizing themes. On anxiety, many students did raise concerns, but for the most part these seem to have been in anticipation of the assessment and, on reflection, many students realised their fears were unfounded. This fits with Henderson, Lloyd and Scott's (2002) finding that looking back on oral assessment can reveal its value and it supports Huxham et al.'s (2012) contention that anxiety may be tied to novelty rather than the nature of the assessment. We would agree with Sparfeldt et al. (2013) that students should have more 'simulations' of oral assessments before they are used for higher stakes. Indeed, the comments made by our participants on what we called 'internal authenticity' resonate with this view: small

scale oral assessments encountered regularly throughout the course may help prepare students for later high stakes presentations.

In other areas there was less resonance with the literature. The students were clearly interested in the issue of fairness, but many of the strongest accusations leveled at oral examinations were not featured in the comments on this implementation. Some were concerned about the comparability of the process between different students: there were different assessors and they asked different contingent questions. It was noted, however, that some existing assessments were also graded by different tutors. Moreover, other students saw the individual nature of the assessments as an advantage: tutors could explore personal understanding and provide indications of areas which might need further work. Some students did mention the possibility of bias and other forms of unfairness, but recording each assessment was seen by most as a guarantee against these problems.

The characteristics of the assessment, however, focused on the understanding engendered by the whole process and how well it was evaluated, the reactiveness of the situation and its authenticity. We will suggest that these characteristics indicate that this type of oral performance assessment might straddle the 'assessment for' and 'assessment of' learning divide.

Black et al. (2004) highlighted the distinction between these two categories of assessment:

> Assessment for learning is any assessment for which the first priority in its design and practice is to serve the purpose of promoting pupils' learning. It thus differs from assessment designed primarily to serve the purposes of accountability, or of ranking, or of certifying competence. (p10)

Within much of the literature, the phrase 'assessment for learning' tends to be focused on the use of assessment to feed in to later teaching and learning: a teacher may

use the responses from pupils to diagnose some misconceptions and design an intervention to address them, or a pupil may use feedback on a piece of assessment to re-examine their understanding of a topic. However, one might also argue that rather than assessment affecting the learning which takes place afterwards, it can also affect the learning which takes place beforehand.

While this 'backwash effect' (Biggs 1999) is often commented on when it has a negative effect (see Corrina's comment about memorizing for examinations), it may be that other forms of assessment might have a positive backwash. This phenomenon of seeing assessment preparation (and even the act of being assessed) as learning opportunities was called 'assessment as learning' by Earl (2003), but unfortunately the term has also been used to describe students' 'procedural compliance' to the sole purpose of passing assessments (Torrance 2007). However, we can include the idea of an assessment's (positive) influence on prior learning under the heading of 'assessment for learning'.

However, by including the notion of 'first priority' within their definition, one might see Black et al. as setting an arbitrary disjunction between 'assessment for learning' and 'assessment of learning'. Certainly within the interviews, there is a sense that the students talk of 'formative' and 'summative' assessments as if they were necessarily disjoint (see, for example, Marie on the idea of authenticity). However, Bennett (2009) notes that such a disjunction is not necessary, nor always helpful: the same act of assessment might simultaneously help the teacher redesign future instruction, help the pupils appreciate the areas on which they need further work and might contribute marks towards some certification. At the crudest level, even a grade or mark is feedback which at least informs the teacher and student about whether there may be a need for further work on a topic.

One might consider, for example, the weekly exercise sheets through both 'assessment for learning' and 'assessment of learning' lenses. Exercise sheets contribute a small amount to the final module mark (and therefore to the summary mark for the year). However, unlike closed book examinations, they are returned to students and tutors are expected to indicate areas of difficulty and provide feedback beyond the marks. We can see some evidence, however, that they do not align fully with the 'assessment for learning' criteria for some students: Sarah's comments about the reactiveness of the oral performance assessments contrasted with her view of the weekly exercise sheets. Her comment that they are returned 'two months later' may exaggerate a normal two-week turnaround, but the issue of the lack of detail and the time delay was present in many students' comments.

However, the students did generally seem to agree that the oral performance assessment provides tutors with very good insight into their understanding of the material. That is, these forms of assessment are candidates for 'assessment of learning' at least on a par with the weekly exercise sheets. But the students' comments appear to suggest that they are also fulfilling the key characteristics of 'assessments for learning'. They provide immediate feedback: tutors can point (either directly or via contingent questioning) to areas of misunderstanding. The act of presenting answers can help solidify understanding and they can have a positive 'backwash' effect: students suggested that being aware they can be asked probing questions led them to focus their preparation directly on developing more relational forms of understanding. Angelina and Tom focused on generalisations and underlying theory, while Ruby distinguished key ideas from 'low steps' in her answers.

Of course, our research must also acknowledge Joughin's (2010) concern about the importance of understanding the context of research, which we described in the

introduction. This was an exploratory study at a research-intensive UK university with students who had very high entry grades. Given that oral performance assessments are commonplace in many other countries and that lack of familiarity appeared related to some of our findings, we would not expect our results to generalize easily outside the UK and it may also be that weaker students have quite different views. However, in a context in which mathematics assessment is dominated by closed book examinations and where there is a push towards diversification, it is important to look carefully at alternatives and to evaluate them in context.

Some concerns about implementing oral performance assessments emerge from this study, but from the students' perspective it appears that some simple solutions can overcome some of these concerns. Some of the issues of fairness can be addressed through the use of video and others need to be balanced against the improved ability of tutors to evaluate understanding. The issue of anxiety may be overcome with more experience (which may have the added bonus of supporting the development of skills for employment and decreasing anxiety with later higher stakes presentation based assessments). It may be that oral performance assessments, implemented carefully, can straddle the 'assessment of learning' and 'assessment for learning' divide and leaven the existing homogenous assessment diet.

Acknowledgment

We would like to thank the MSOR Network in the UK for funding this project through the Mathematical Sciences HE Curriculum Innovation Project, part of the National HE STEM Programme.

References

Attride-Stirling, J. 2001. 'Thematic networks: an analytic tool for qualitative research.' *Qualitative Research* 1(3): 385-405.

Bennett, R. 2009. *A critical look at the meaning and basis of formative assessment.* Princeton, NJ: Educational Testing Service.

Biggs, J. 1999. *Teaching for quality learning at university*. Buckingham: SRHE and Open University Press.

Biglan, A. 1973. 'The characteristics of subject matter in different academic areas.' *Journal of Applied Psychology* 57(3): 195-203.

Birenbaum, M, K. Brueur, E. Cascallar, F. Dochy., Y. Dori, J. Ridgway, R. Wiesemes and G. Nickmans . 2006. 'A learning integrated assessment system.' *Educational Research Review* 1(1): 61-67.

Black, P., C. Harrison, C. Lee, B. Marshall and D. Wiliam. 2004. 'Working inside the black box: assessment for learning in the classroom.' *Phi Delta Kappan* 86(1): 8-21.

Briggs, D. 1980. 'A study of the influence of handwriting upon grades using examination scripts.' *Educational Review 32*(2): 186-193.

Brown, G., J. Bull and M. Pendlebury. 2013. *Assessing student learning in higher education*. Abingdon: Routledge.

Brown, S. and P. Knight. 1994. *Assessing learners in higher education*. Abingdon: Routledge.

Davis, M. and I. Karunathilake. 2005. 'The place of the oral examination in today's assessment systems.' *Medical Teacher 27*(4): 294-297.

De Vita G.. and P.Case. 2003. 'Rethinking the internationalisation agenda in UK higher education.' *Journal of Further and Higher Education* 27(4): 383-398.

Diaz, R., C.Glass, D. Arnkoff , and M. Tanofsky-Kraff. 2001. 'Cognition, anxiety, and prediction of performance in 1st-year law students.' *Journal of Educational Psychology 93*(2): 420-429.

Earl, L. 2003. *Assessment as learning.* Thousand Oaks: Corwin Press.

Gillham, B.. 2005. *Research interviewing: The range of techniques: a practical guide*. Maidenhead: McGraw-Hill International.

Henderson, J., Peter Lloyd and Heather Scott,. 2002. ''In the real world we're all put on the spot at some time or other, so you need to be prepared for it': an exploratory study of an oral method of assessing knowledge of mental health law.' *Social work education 21*(1): 91-103.

Heyneman, S. P. 2004. 'Education and corruption.' *International Journal of Educational Development 24*(6): 637-648.

Huxham, M., F. Campbell and J.Westwood. 2012. 'Oral versus written assessments: a test of student performance and attitudes.' *Assessment & Evaluation in Higher Education* 37(1): 125-136.

Iannone, P. and A. Simpson. eds. 2012a. *Mapping University Mathematics Assessment Practices*. Norwich, UK: University of East Anglia.

Iannone, P. and A. Simpson. 2012b. 'Oral assessment in mathematics: implementation and outcomes.' *Teaching Mathematics and its Applications 31*(4): 179-190.

Iannone, P. and A. Simpson. 2014. 'Students' preferences in undergraduate mathematics assessment.' *Studies in Higher Education* (ahead-of-print), 1-22.

Joughin, G. 1998. 'Dimensions of oral assessment.' *Assessment & Evaluation in Higher Education* 23(4): 367–78.

Joughin, G. 2010. 'The hidden curriculum revisited: a critical review of research into the influence of summative assessment on learning.' *Assessment & Evaluation in Higher Education 35*(3): 335-345.

LMS (London Mathematical Society) (2010) *Mathematics degrees, their teaching and assessment*. London Mathematical Society Position Paper, http://www.lms.ac.uk/sites/lms.ac.uk/files/Mathematics/Policy_repors/2010%20teaching_position_statement.pdf accessed June 2014.

Norton, L., B. Norton and L. Shannon. 2013. 'Revitalising assessment design: what is holding new lecturers back?' *Higher Education* 66(2): 233–251.

Skemp, R. 1976. 'Instrumental understanding and relational understanding.' *Mathematics Teaching 77*: 20-26.

Sparfeldt J, D. Rost, U. Baumeister and O. Christ. 2013. 'Test anxiety in written and oral examinations.' *Learning and Individual Differences 24*: 198-203.

Steen, L. ed. 2006. *Supporting Assessment in Undergraduate Mathematics*. Washington, DC.:The Mathematical Association of America.

Torrance, H. 2007. 'Assessment as learning? How the use of explicit learning objectives, assessment criteria and feedback in post-secondary education and training can come to dominate learning.' *Assessment in Education 14*(3): 281-294.

Yorke, M. 2011. 'Summative assessment: dealing with the 'measurement fallacy'.' *Studies in Higher Education 36*(3): 251-273.

A: Prove that if a graph has at least 11 vertices then either it or its complement must be non-planar.

B: Show that every connected planar graph with less than 12 vertices has a vertex of degree 4 or less. [Hint: argue by contradiction to get a lower bound for the number of edges which contradicts the upper bound which follows from Euler's formula].

C: For each graph find the minimum spanning tree and show it is unique:

(a) Q_3 with the usual binary vertex label and weight $w(ij)=i+j;$

(b) K_5 with vertices {1, … , 5} and weight $w(ij)=i+j^2$ where $i < j$.

D: Draw all forests of 5 vertices and justify your answers.

Table 1. The questions set for the oral performance assessment.

Grade	Solution	Key ideas and application	Clarity and explanation
5	Complete solution outline given with no extra help needed	Clearly identified key ideas behind the problem and shown how they apply elsewhere	Explains clearly and concisely, even in unfamiliar areas
4	Complete solution given with some extra help	Identified key ideas or shown how solution approach might apply elsewhere	Explains clearly and concisely in prepared areas and generally clear elsewhere
3	Complete solution given with substantial extra help	Has identified some key ideas, but may not fully distinguish key ideas from calculations or details OR shown some sense of wider application of solution	Explanations need a little probing to clarify
2	Complete solution not obtained, but some key steps made without help	Does not have key ideas or any sense of wider application	Explanations need to be drawn out at length
1	Complete solution not obtained, but some key steps made with help	Does not have key ideas or any sense of wider application	Has difficulty giving any explanations

Table 2. Marking matrix agreed by the assessors.

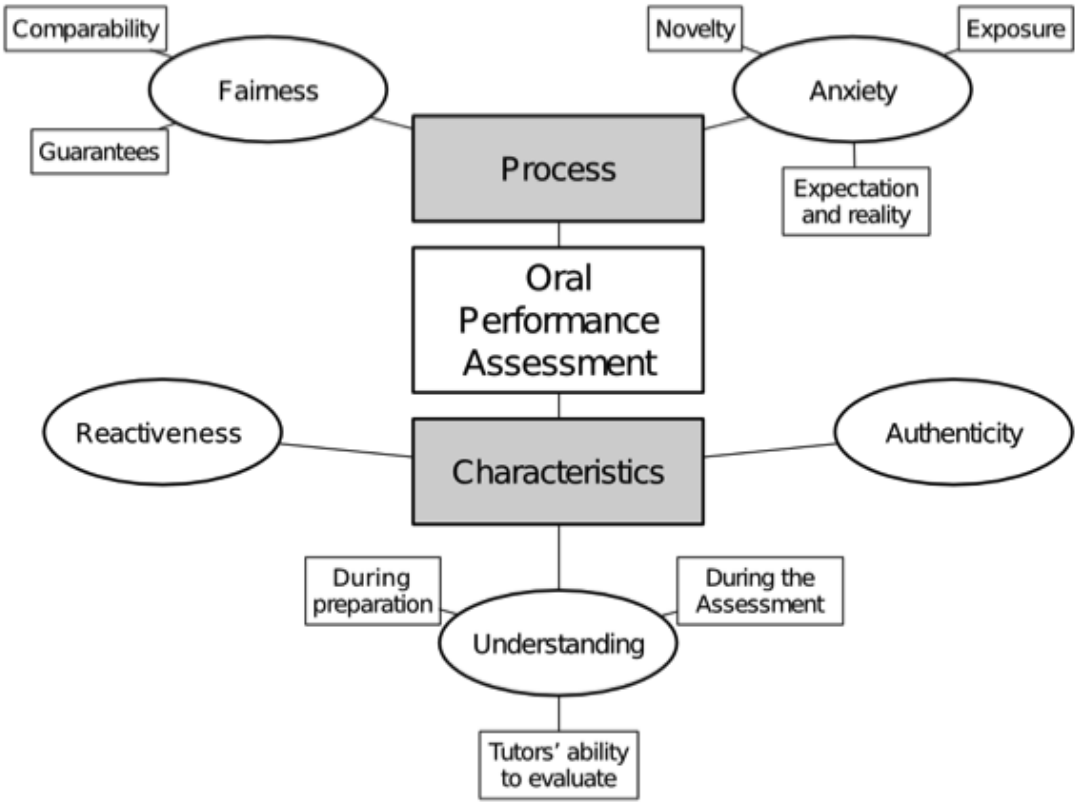

Figure 1. Resultant thematic network.

Figure Captions

Table 1. The questions set for the oral performance assessment.

Table 2. Marking matrix agreed by the assessors.

Figure 1. Resultant thematic network.